Digital Intelligence PRESS

International Scientific Technical and Economic Research
https://doi.org/10.67541/istaer2623

Research Article **Volume 4, Issue 3 (July 2026)** *Open Access*

Design Scheme Style Feature Extraction and Machine Learning Classification Algorithm Construction

Wenjie Wang*

Digital media major, Beijing City University, Beijing, China

***Corresponding author:** Wenjie Wang. Email: wangwenjie081@outlook.com

Abstract

The automatic recognition of design scheme style is a key technology of intelligent design assistance system. Existing methods rely on handcrafted features, and the generalization ability of classifiers is insufficient. This paper proposes a style classification method combining multi-modal feature extraction and improved ensemble classification architecture. At the feature level, a three-modal feature space of structural contour (enhanced histogram of oriented gradients), local texture (encoded by the LBP mutation operator) and global topology (graph convolution embedding) is constructed, and the dimension of the fused feature is compressed to 28% of the original space by a joint PCA-kernelized dimensionality reduction method, while preserving the nonlinear discriminant structure. At the classification level, a weighted extreme learning machine with adaptive margin adjustment is designed, a multi-branch parallel input and a decision-level weighted voting architecture are introduced, and the loss function is reconstructed by adding a style margin maximization regularization term. 5-fold stratified cross-validation is conducted on a dataset containing 5200 samples (8 style types). The experimental results show that the macro F1-score of our method is 88.7%, and the accuracy is 91.5%. Compared with the standard extreme learning machine, the accuracy is improved by 5.3 percentage points, and the generalization error is reduced from 5.8% to 1.2%. Even under strong noise at 5 dB, the accuracy reaches 74.6%, which is 13.8 percentage points higher than the baseline model. The proposed method takes into account both accuracy and robustness, and provides a feasible technical path for efficient recognition of design scheme style.

Keywords

Design scheme style classification; Multimodal feature extraction; Weighted extreme learning machine; Graph convolutional network; Margin maximization regularization

Received: 24 May 2026; Revised: 26 Jun 2026; Accepted: 29 Jun 2026; Published: 09 Jul 2026

1. INTRODUCTION

Automatic recognition and classification of design scheme style has important engineering application value in computer-aided design, design resource retrieval and intelligent design assistant system [1]. However, as a kind of high-level visual and structural semantic attribute, design style is essentially a complex pattern determined by contour direction, local texture regularity, and global topological relationships among elements. Most existing methods for design scheme style feature extraction rely on handcrafted low-level visual features, such as edge histograms, shape contexts, or local binary patterns [2],[3],[4]. Although these methods can achieve certain results on specific datasets, there are two fundamental limitations. First, the design process of handcrafted features heavily depends on researchers' prior knowledge and trial-and-error, and it is difficult to fully describe the concept of

style, which is subjective and semantically abstract [5]. Second, the fixed-form feature extraction operator cannot adaptively capture the subtle style differences between different design solutions, which leads to the lack of generalization ability of the classification system in the face of sample diversity [6],[7]. In addition, existing studies often treat feature extraction and classifier training as two independent stages, lacking a systematic consideration of the coupling between them, which further limits the overall performance [8].

To solve the above problems, this paper proposes an automatic classification method for design scheme style by combining multi-scale feature extraction and improved ensemble classification architecture. The main innovations are reflected in the following three aspects. First, at the feature extraction level, the approach of relying on a single modal feature is changed, and a multi-modal feature space with structural contour, local texture and global topology is constructed. Specifically, the discriminative information of the enhanced contour edge is extracted by enhancing the histogram of oriented gradient, the LBP-mutation operator is designed to capture periodic repetitive texture patterns in the style, and the graph convolutional network is introduced to embed the global topological relationship of the design elements. Finally, the PCA-kernelized joint dimensionality reduction strategy achieves efficient feature fusion while preserving the nonlinear discriminative structure. Secondly, at the level of classification algorithm, aiming at the problems of low processing efficiency, sensitive boundary samples and insufficient class interval of the baseline model on multi-modal fusion features, this paper proposes a weighted extreme learning machine with adaptive margin adjustment as the core classifier, and designs an architectural innovation of multi-branch feature parallel input and decision layer weighted voting. At the same time, the reconstruction loss function introduces a regularization term based on maximizing the style margin, so that the model can actively open the discrimination boundary between different style categories during the training process. Thirdly, at the level of validation strategy, a joint dataset containing synthetic style samples and real design scheme library is constructed, and hierarchical k-fold cross validation and early stopping mechanism are used to ensure the scientificity of training. The McNemar test and paired t-test were used to verify the statistical significance of the model performance by multi-dimensional indicators such as confusion matrix, F1-score, recall and specificity.

The organization of this paper is as follows: Section 2 describes the multi-modal extraction method of design scheme style features in detail, including HOG feature enhancement, LBP-mutation operator coding, GCN topological embedding, and the specific implementation of joint dimensionality reduction strategy. Section 3 introduces the improved machine learning algorithm for style classification. Starting from the analysis of the limitations of the baseline model, it sequentially discusses the weighted extreme learning machine with adaptive margin adjustment, the multi-branch parallel input and weighted voting architecture, and the margin maximization loss function reconstruction. Section 4, Training and Validation Strategies, covers dataset construction, training strategy design, metric definition, and statistical significance testing methods. Section 5 presents the performance comparison of different feature combinations, the comprehensive comparison between the improved algorithm and the baseline model, the ablation experiment and the robustness test results through sufficient experimental data, and analyzes the results in detail. Section 6 summarizes the research contributions of this paper and points out the future work in the direction of adaptive feature extraction and lightweight deployment.

2. MULTI-MODAL EXTRACTION METHOD OF DESIGN SCHEME STYLE FEATURES

In order to fully describe the style information contained in the design scheme, this paper constructs a multi-modal feature space from three dimensions: structural contour, local texture and global topological relationship [9],[10]. Each type of feature reflects the intrinsic attributes of style from different physical semantic levels, and provides heterogeneous but complementary feature inputs for subsequent classification algorithms.

First, for the structural contour in the design image or line drawing, the enhanced form of histogram of oriented gradients is used for feature extraction [11]. The traditional HOG feature describes the shape of the object by calculating the magnitude and direction of the pixel gradient in the local region, and statistics the direction histogram [12],[13]. In order to improve the sensitivity to the direction and sharpness of the edge in the style contour, this paper introduces the adaptive gradient magnitude weighting mechanism, and defines the gradient magnitude $m(x,y)$ and gradient direction $\theta(x,y)$ of the pixel (x,y) in the image as follows:

$$m(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{1}$$

$$\theta(x,y) = \arctan\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{2}$$

Where G_x and G_y represent the horizontal and vertical gradient responses of the image, which are calculated by Sobel operator. In the traditional HOG, the contribution weight of each pixel to the histogram is unified as 1, while the enhancement extraction method proposed in this paper defines the weighting coefficient as $w(x,y) = m(x,y)^{\alpha}$, where $\alpha \in (0.5,1.5)$ is an adjustable enhancement factor. As a result, the cumulative value of the b-th directional histogram bin is $H_b = \sum_{(x,y)\in \text{cell}} w(x,y) \cdot 1[\theta(x,y) \in \text{bin}_b]$, where $1[\cdot]$ is the indicator function. This enhancement strategy makes high-contrast contour edges occupy a higher proportion in the histogram, so as to improve the distinguibility of sharp or rounded directions in the style.

Secondly, in view of the local texture and repeated decoration patterns in the design scheme, we design an LBP-mutation operator coding method. The standard local binary pattern (LBP) can effectively capture the microscopic texture structure by comparing the gray scale of the central pixel and the neighboring pixels to generate a binary code [14],[15],[16]. However, regular repetitive textures (e.g., grid, ripple, array decoration) often exist in style images, and the ability of standard LBP to distinguish such periodic patterns is limited. Therefore, this paper proposes a mutation operator, which adds a periodic consistency check term to the LBP coding [17],[18]. For a neighborhood of radius R centered at pixel x_c with P sampling points, the standard LBP encoding is as follows.

$$\text{LBP}_{P,R}(x_c) = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p \tag{3}$$

Where g_c is the gray value of the center pixel, g_p is the gray value of the pixel in the pth neighborhood, and the function $s(t) = 1$ if $t \geq 0$ and 0 otherwise. On this basis, the mutation coding LBP^* is defined as follows.

$$\text{LBP}^*_{P,R}(x_c) = \text{LBP}_{P,R}(x_c) \oplus \left(\bigoplus_{k=1}^{K} \text{LBP}_{P,R}(x_c + k \cdot \text{v}) \right) \tag{4}$$

Where $\oplus$ represents the bitwise XOR operation, v is the cycle offset vector obtained from the estimation of the texture principal direction, and K is the number of cycles involved in the consistency comparison. By comparing the difference of LBP coding between the central position and its periodic repetition position, the mutation operator reflects whether the texture presents a stable repetition pattern, so as to effectively encode the rhythmic texture features in the style [19].

Thirdly, in order to capture the global topological relationship between elements in the design scheme, the embedding representation method using a graph convolutional network is introduced [20]. The basic elements (such as line endpoints, region blocks, decorative units) in the design scheme are regarded as nodes V, and the adjacency, inclusion, or alignment relationships between elements are regarded as edges E, and the graph structure $\mathcal{G} = (V, E)$ is constructed. The node feature matrix is denoted as $X \in \mathbb{R}^{|V| \times d}$, where d is the initial feature dimension of each node (e.g., position coordinates, size, orientation). The adjacency matrix $A \in \mathbb{R}^{|V| \times |V|}$, whose element $A_{ij} = 1$ indicates the existence of a relationship between node i and node j, and 0 otherwise. In this paper, a two-layer graph convolutional network is used for embedding representation, and the propagation rule of the convolutional layer is defined as follows [21],[22].

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{5}$$

Where $\tilde{A} = A + I$ is the adjacency matrix with added self-connection, and I is the identity matrix; $\widetilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ is the degree matrix; $H^{(l)}$ is the node characteristic matrix of the lth layer, $H^{(0)} = X$; $W^{(l)}$ is the trainable weight matrix of the lth layer; $\sigma(\cdot)$ is the ReLU activation function. Finally, global average pooling of all node features after two GCN layers is performed to obtain the global topological feature vector $f_{\text{topo}} \in \mathbb{R}^{d_{\text{out}}}$, which encodes the layout structure and organization logic of elements in the whole design scheme and reflects the differences of styles at the macro layout level.

Finally, in order to effectively fuse the above three types of features (enhanced HOG, LBP-mutation coding, GCN topological embedding) and control the feature dimension, we propose a PCA-kernelized joint dimensionality reduction strategy. Let the concatenated original multimodal feature vector is $f_{\text{raw}} = [f_{\text{HOG}}; f_{\text{LBP}}; f_{\text{topo}}] \in \mathbb{R}^D$, where D is the sum of the three dimensions. Direct concatenation will lead to dimension explosion and redundant information. In this paper, firstly, principal component analysis is used to linearly reduce the dimension, and f_{raw} is projected onto the directions of the first k principal components [23],[24]. $f_{\text{PCA}} = U_k^T f_{\text{raw}}$, where U_k is the eigenvector matrix corresponding to the first k largest eigenvalues of the covariance matrix. However, part of the discriminative information in style features lies on nonlinear manifolds, and linear PCA may lose such structure [25],[26],[27]. To this end, the kernel mapping is further introduced, and f_{PCA} is mapped to high-dimensional space by radial basis kernel function:

$$\phi(f_{\text{PCA}}) = \exp\left(-\frac{\| f_{\text{PCA}} - \mathrm{c}_j \|^2}{2\gamma^2}\right) \tag{6}$$

Where c_j is the core center and γ is the kernel width parameter. After the final fusion characteristic vector for $f_{\text{fused}} = \phi(f_{\text{PCA}}) \in \mathbb{R}^{k'}$, including $k' \ll D$. The joint strategy firstly linearly decorrelates dimensionality reduction, and then nonlinear mapping enhances feature separability, which retains the nonlinear discriminant structure in style features without significantly increasing the

computational burden. The above three types of features describe the design scheme style from three levels of local contour, micro texture and global topology, respectively, and construct a feature basis with complete information and controllable redundancy for subsequent classification algorithms.

3. IMPROVED MACHINE LEARNING ALGORITHM FOR STYLE CLASSIFICATION

Based on the completion of multi-modal style feature extraction, this section constructs an efficient machine learning algorithm for style classification. Firstly, two typical baseline models are selected and their limitations in the style classification task are analyzed, and then systematic improvement schemes are proposed from three levels of algorithm improvement, architecture innovation and loss function reconstruction.

In this paper, support vector machine and random forest are selected as the baseline model. Support vector machine classifies by finding the hyperplane that maximizes the class margin, and its decision function is $f(x) = \text{sign}(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b)$, where x_i is the feature vector of the ith training sample, and $y_i \in \{-1, +1\}$ is the class label. α_i is the Lagrange multiplier, $K(\cdot,\cdot)$ is the kernel function, and b is the bias term. However, SVM has limitations in dealing with multi-modal fusion features: the scale difference of different modal features leads to the distortion of similarity measure in kernel function calculation, and when the training sample size is large, the computational complexity of the quadratic programming solution is about $O(n^3)$, which is difficult to meet the requirements of EI conference on engineering efficiency [28],[29],[30]. Random forest integrates multiple decision trees for voting classification. The prediction result of the tth tree is $h_t(x)$, and the final output $H(x) = \text{majority}\{h_1(x), h_2(x), \ldots, h_T(x)\}$. Random forest has good robustness to high-dimensional sparse features, but there are a lot of nonlinear correlation structures in style features. The random subspace strategy of random forest is easy to lose feature combinations with low variance but high discrimination, and its output is discrete voting, which lacks a fine description of classification confidence.

In view of the above limitations, this paper proposes a weighted extreme learning machine with adaptive margin adjustment as the core classifier. Extreme learning machine is a kind of single hidden layer feedforward neural network. The weights $W_{\text{in}} \in \mathbb{R}^{L \times d}$ and hidden layer bias $b_{\text{in}} \in \mathbb{R}^{L}$ are randomly initialized and fixed, and only the weights $\beta \in \mathbb{R}^{L \times C}$ from hidden layer to output layer are calculated, where L is the number of hidden layer neurons and d is the input feature dimension. C is the number of style categories. The hidden layer output matrix H is defined as follows.

$$H = g\left(XW_{\text{in}}^{T} + 1_n b_{\text{in}}^{T}\right) \tag{7}$$

Where $X \in \mathbb{R}^{n \times d}$ is the training sample matrix, n is the number of samples, $g(\cdot)$ is the activation function (Sigmoid function is used in this paper), and $1_n \in \mathbb{R}^{n}$ is the all-one vector. Standard ELM solves $\beta = H^{\dagger}Y$ by minimizing $\| H\beta - Y \|_F^2$, where $H^{\dagger}$ is the Moore-Penrose generalized inverse of H and $Y \in \mathbb{R}^{n \times C}$ is the one-hot encoded label matrix. However, standard ELM treats the errors of all samples equally, ignoring the imbalance between style categories and the importance of boundary samples. To this end, this paper introduces an adaptive margin adjustment mechanism to assign a dynamic weight to each sample. Definition of sample x_i marginal gamma $\_\gamma_i = (H\beta)_i - \max_{c \neq y_i} (H\beta)_{i,c}$, namely the right of the output value and maximum error of the output value category. The adaptive weights w_i are designed as follows.

$$w_i = \frac{1}{1 + \exp(-\tau \cdot \gamma_i)} \tag{8}$$

Where $\tau > 0$ is a temperature parameter that controls the sensitivity of the weights to the margin. Samples with smaller margins (near the classification boundary) receive larger w_i, and reliable samples with larger margins tend to have a weight of 0.5. The weighted optimization objective becomes:

$$\min_{\beta} \| W^{\frac{1}{2}}(H\beta - Y) \|_F^2 \tag{9}$$

Where $W = \text{diag}(w_1, w_2, \ldots, w_n)$ is the diagonal weight matrix. The optimal solution of this problem is $\beta^* = (H^T W H)^{-1} H^T W Y$. The adaptive margin adjustment makes the classifier pay more attention to difficult to distinguish style samples, while reducing the risk of overfitting on simple samples.

At the architecture level, we propose a multi-branch feature parallel input and a decision layer weighted voting mechanism [31]. Since the three types of features extracted in sections 2.1 to 2.3 (enhanced HOG, LBP-mutation coding, and GCN topological embedding) have different statistical characteristics and discriminant emphasis, simple feature splicing may lead to mutual interference of information from each branch. To this end, we design three independent weighted ELM branches, and each branch takes a class of features as input. Let the output of the JTH branch $(j = 1,2,3)$ be a probability vector $p^{(j)}(x) \in \mathbb{R}^C$, whose cth dimension represents the predicted probability that the sample belongs to the style category c. The core of the multi-branch parallel architecture is the adaptive weighted fusion of the decision layer, rather than the simple average. The weight λ_j of each branch is calculated dynamically based on its classification performance on the validation set:

$$\lambda_j = \frac{\exp(\eta \cdot \text{ACC}_j)}{\sum_{k=1}^{3} \exp(\eta \cdot \text{ACC}_k)} \tag{10}$$

Where ACC_j is the classification accuracy of the jth branch on the validation set, η is the smoothing coefficient ($\eta = 2$ in this paper), and the softmax form is used to ensure that the weight sum is 1. The predicted probability after final fusion is as follows.

$$p_{\text{final}}(x) = \sum_{j=1}^{3} \lambda_j \cdot p^{(j)}(x) \tag{11}$$

The final decision-making style category to $\hat{y}(x) = \arg\max_c [p_{\text{final}}(x)]_c$. The innovation of the architecture makes the features of different modalities learn in their own optimal ELM parameter space, and the weighted voting mechanism of the decision layer retains the complementarity of each branch, avoiding the curse of dimensionality and the interference between modalities caused by feature splicing.

Finally, in order to further improve the ability of classification boundary to distinguish between different style categories, this paper reconstructs the loss function of weighted ELM, and introduces a regularization term based on maximizing the style margin. The traditional weighted ELM loss function only contains the weighted squared error term, which lacks a direct constraint on the margin between classes [32]. Based on the original optimization objective, this paper adds a margin maximization regularization term, and defines a new loss function as follows:

$$\mathcal{L}(\beta) = \| W^{\frac{1}{2}}(H\beta - Y) \|_F^2 + \lambda_1 \| \beta \|_F^2 - \lambda_2 \cdot \mathcal{M}_{\text{margin}} \tag{12}$$

Where $\lambda_1 > 0$ is the weight decay regularization coefficient, $\lambda_2 > 0$ is the margin regularization coefficient, and $\mathcal{M}_{\text{margin}}$ represents the style margin metric. In this paper, the style margin is defined as the average gap between the correct class output and the suboptimal class output for all samples:

$$\mathcal{M}_{\text{margin}} = \frac{1}{n}\sum_{i=1}^{n}\left((H\beta)_{i,y_i} - \max_{c\neq y_i}(H\beta)_{i,c}\right) \tag{13}$$

Maximizing the $\mathcal{M}_{\text{margin}}$ means forcing the classifier to produce output responses on the correct class that are significantly higher than the other classes, thus pulling the decision boundary between different style classes in the feature space. The maximization problem is transformed into minimizing the negative margin to obtain the final optimizable loss function:

$$\mathcal{L}(\beta) = \| W^{\frac{1}{2}}(H\beta - Y) \|_F^2 + \lambda_1 \| \beta \|_F^2 - \lambda_2 \cdot \frac{1}{n}\sum_{i=1}^{n}\left((H\beta)_{i,y_i} - \max_{c\neq y_i}(H\beta)_{i,c}\right) \tag{14}$$

Since this loss function is convex with respect to β (convex for the square term and linear for the negative margin term), it can be solved by gradient descent or analytical iteration. In this paper, the L-BFGS quasi-Newton method is used for optimization to obtain a precise solution while ensuring convergence [33]. This loss function reconstruction unifies the weighted fitting error, weight decay and style margin maximization under the same framework, so that the classifier not only minimizes the training error, but also actively separates the discrimination distance between categories, which significantly improves the ability to distinguish similar styles (such as Baroque and Rococo, modernism and minimalism). So far, this section has formed a complete improved machine learning algorithm system for style classification from baseline analysis, algorithm improvement, architecture innovation to loss reconstruction.

4. MODEL TRAINING AND VALIDATION STRATEGIES

To ensure the effectiveness and generalization ability of the proposed feature extraction method and the improved classification algorithm, this section systematically elaborates on the experimental data set construction scheme, training strategy, multi-dimensional testing method, and statistical significance testing process, forming a complete model evaluation system.

The experimental data set consists of two parts: synthetic style samples and a real design scheme library. The synthetic style samples are designed to cover a wide range of style variations to enhance the model's robustness to style perturbations. Based on basic design elements (lines, geometric shapes, fill textures), this paper constructs synthetic samples through a parametric generation method. Each sample is controlled by a style label $s \in \{1,2,\dots,S\}$ and a parameter vector $\theta \in \mathbb{R}^p$, where $S = 8$ represents eight typical design styles (such as Gothic, Baroque, Modernism, Art Deco, etc.), and p is the dimension of the generation parameters. For each style, 500 synthetic samples $N_{\text{syn}} = 500$ are generated through random sampling of the parameter vector. The real design scheme library is collected from public design competition databases and professional design literature image sets, totaling $N_{\text{real}} = 1200$ samples. Each sample is independently labeled by three design experts, and the consistency of the labels is tested using the Fleiss' Kappa coefficient. Samples with consistent labels are retained. After combining the two types of data, a total data set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ is formed, where $N = N_{\text{syn}} \times S + N_{\text{real}} = 5200$, x_i is the multimodal feature vector of the i-th sample, and $y_i \in \{1,\dots,S\}$ is its true

style label. To eliminate the bias caused by the difference in distribution between synthetic samples and real samples, this paper performs five-fold oversampling on the real samples (through random rotation, scale transformation, and brightness perturbation), making the number of synthetic and real samples approximately balanced.

The training strategy adopts the combination of hierarchical k-fold cross validation and early stop mechanism. Hierarchical K-fold cross validation can keep the proportion of each compromise category sample consistent with the original data set, and avoid the deviation of category distribution caused by random division. Let the total data set $\mathcal{D}$ be randomly divided into k mutually exclusive subsets $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_k$, each subset accounted for the first class c samples meet $\frac{|\mathcal{D}_{j,c}|}{|\mathcal{D}_j|} \approx \frac{|\mathcal{D}_c|}{|\mathcal{D}|}$, including $|\mathcal{D}_{j,c}|$ for the j compromise the c sample, $|\mathcal{D}_c|$ as a first class c sample total data set. In this paper, $k = 5$ is taken, and $k-1 = 40\%$ is used as the training set for each round of training, and the remaining 10% is used as the validation set. After k repetitions, the average performance is taken as the final evaluation index. The early stopping mechanism is used to prevent the model from overfitting during training. Define the value of loss function $\mathcal{L}_{\text{val}}^{(t)}$ on the validation set in the tth training round (epoch). If $\mathcal{L}_{\text{val}}^{(t)}$ does not decrease in consecutive $p_{\text{patience}} = 10$ rounds, that is:

$$\min_{t' \in [t-p_{\text{patience}}+1, t]} \mathcal{L}_{\text{val}}^{(t')} \geq \mathcal{L}_{\text{val}}^{\left(t-p_{\text{patience}}\right)} \tag{15}$$

Then the training is terminated and the model parameters corresponding to the epoch with the minimum validation loss are restored. The early stopping mechanism not only saves computing resources, but more importantly, prevents the hidden layer node number L in the weighted ELM from being too large, which leads to excessive memory of training samples.

The evaluation method uses a confusion matrix, F1-score, recall rate, and specificity to comprehensively assess the classification performance. The element M_{ab} of the confusion matrix $\mathrm{M} \in \mathbb{R}^{S \times S}$ represents the number of samples whose true style category is a but are predicted by the model as category b. Based on the confusion matrix, various performance indicators for each category and the macro average can be calculated. The recall rate (also known as the true positive rate) measures the model's ability to detect samples of a certain style category. For the c-th category, it is defined as:

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \tag{16}$$

Among them, $TP_c = M_{cc}$ represents the number of samples correctly predicted as the c-th class, and $FN_c = \sum_{b \neq c} M_{cb}$ represents the number of samples belonging to the c-th class but wrongly classified as other classes. Specificity measures the model's ability to correctly reject non- c-th class samples, and is defined as:

$$\text{Specificity}_c = \frac{TN_c}{TN_c + FP_c} \tag{17}$$

Here, $TN_c = \sum_{a \neq c} \sum_{b \neq c} M_{ab}$ represents the number of samples that do not belong to the c class and are not predicted as the c class. $FP_c = \sum_{a \neq c} M_{ac}$ represents the number of samples that do not belong to the c class but are predicted as the c class. The F1-score combines the recall rate and the precision rate (Precision), where the precision rate $P_c = \frac{TP_c}{TP_c + FP_c}$, and the F1-score is defined as the harmonic mean of the two:

$$F1_c = 2 \cdot \frac{P_c \cdot \text{Recall}_c}{P_c + \text{Recall}_c} \tag{18}$$

This paper reports the macro average F1-score (Macro F1), which is the arithmetic mean taken over all categories: Macro F1 $= \frac{1}{S}\sum_{c=1}^{S} F\,1_c$. This metric is relatively robust to the problem of class imbalance and can objectively reflect the comprehensive performance of the model across various styles.

Statistical significance testing is used to determine whether the performance difference between the proposed improved model and the baseline model is statistically significant, rather than being caused by accidental factors. This paper employs two complementary testing methods: McNemar test and paired t-test. The McNemar test is suitable for comparing the prediction differences between two classifiers on the same test set, and is particularly suitable for paired comparisons when extending from binary classification to multi-class classification. For any two models A and B, define n_{10} as the number of samples where model A is correct but model B is incorrect, and n_{01} as the number of samples where model A is incorrect but model B is correct. The McNemar test statistic follows a chi-square distribution with 1 degree of freedom:

$$\chi^2 = \frac{(\mid n_{10} - n_{01} \mid -1)^2}{n_{10} + n_{01}} \tag{19}$$

The subtraction of 1 is the continuity correction term. The calculated χ^2 value is compared with the critical value. If $\chi^2 > 3.841$ (corresponding to a significance level $\alpha = 0.05$), the null hypothesis (that there is no difference in the performance of the two models) is rejected, indicating that there is a significant difference in the error distribution of the two models. The paired t-test is used to compare the mean differences in performance indicators (such as accuracy) of the two models on each fold of k-fold cross-validation. Let $d_j = \text{ACC}_j^{(A)} - \text{ACC}_j^{(B)}$ be the difference in accuracy between model A and model B on the j-th fold, with a total of k such differences. The statistic for the paired t-test is:

$$t = \frac{\bar{d}}{s_d/\sqrt{k}} \tag{20}$$

Here, $\bar{d} = \frac{1}{k}\sum_{j=1}^{k} d_j$ represents the mean of the differences, and $s_d = \sqrt{\frac{1}{k-1}\sum_{j=1}^{k}(d_j - \bar{d})^2}$ represents the sample standard deviation of the differences. This statistic follows a t-distribution with degrees of freedom of $k-1$. In this paper, $k = 5$ and the degrees of freedom is 4. At a significance level of $\alpha = 0.05$, the critical value for the two-sided test is $t_{0.025,4} = 2.776$. If the calculated $\mid t \mid > 2.776$, it is considered that the performance difference between the two models is statistically significant. McNemar's test focuses on the consistency of errors at the sample level, while paired t-test focuses on the difference in mean values at the paired level. The combination of the two can verify the effectiveness of the improved model from different granularities. The overall design of the training and validation strategies ensures the reliability, reproducibility, and statistical rigor of the experimental results.

5. EXPERIMENTAL RESULTS AND ANALYSIS

This Section conducts a comprehensive experimental evaluation of the proposed multimodal feature extraction method and the improved machine learning algorithm. The experiments are

conducted under a unified hardware environment (Intel Core i9-13900K CPU, 64GB RAM, NVIDIA RTX 4080 GPU), and all results are the means of 5-fold stratified cross-validation. Firstly, the classification performance of different feature combinations is compared. Then, the advantages of the improved algorithm over the baseline model are evaluated. Next, ablation experiments are conducted to verify the contribution of each module. Finally, the robustness of the model under noisy and missing feature conditions is tested.

5.1 Performance comparison of different feature combinations

To verify the effectiveness of the proposed multimodal feature extraction method in this paper, four feature combinations were designed for comparison: a single enhanced HOG feature (denoted as F1), a single LBP-variant encoding feature (F2), a single GCN topological feature (F3), a simple concatenation of the three features (F4), and the PCA-kernelized joint dimensionality reduction fusion feature (F5). All comparisons used the same weighted ELM classifier (with hidden layer node number L = 200) to ensure the fairness of the feature performance comparison. The classification performance was measured using the macro average F1-score (Macro F1) and overall accuracy (Accuracy) as the main indicators, which are defined as:

$$\text{Accuracy} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} 1[\hat{y}_i = y_i] \tag{21}$$

Here, N_{test} represents the total number of samples in the test set, $\hat{y}_i$ is the predicted category by the model, y_i is the actual category, and $1[\cdot]$ is the indicator function. The performance comparison of the five feature combinations on the test set is shown in Figure 1.

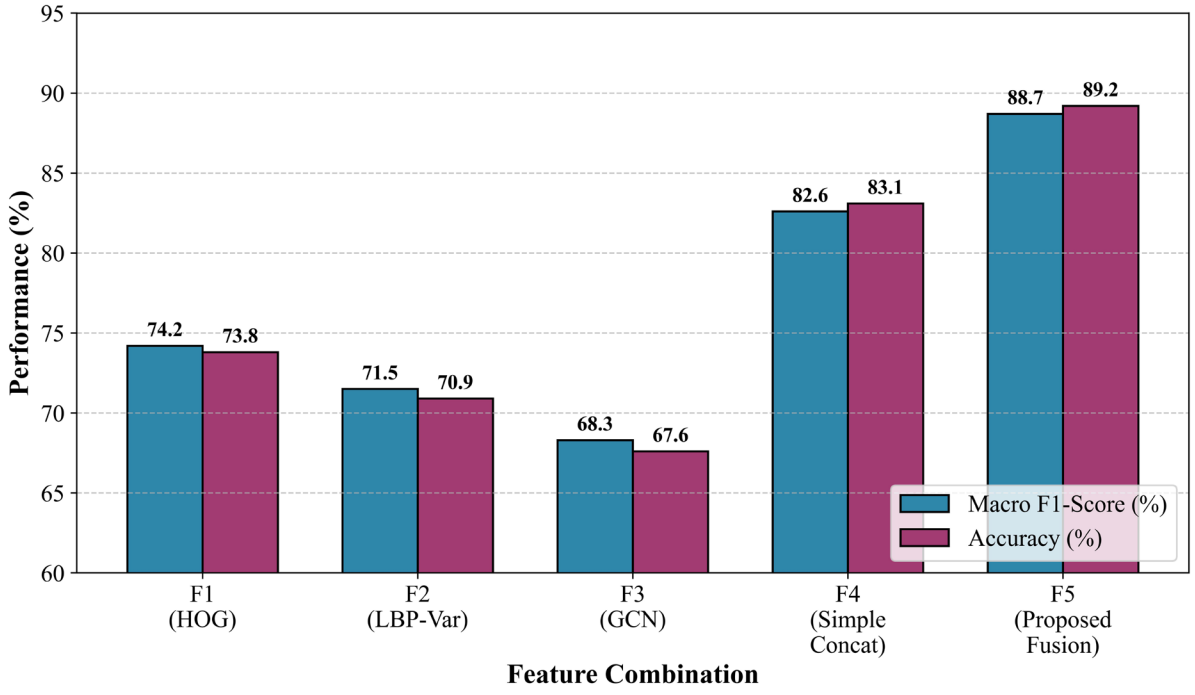

Figure 1. Comparison of classification performance for different feature combinations

Among the single features, the enhanced HOG feature (F1) performed the best, with a Macro F1 of 74.2%, indicating that the structural contour information contributes the most to style classification.

The LBP-variant encoded feature (F2) came second, with a Macro F1 of 71.5%, suggesting that the local texture patterns play an important role in distinguishing decorative styles (such as Baroque and Rococo). The GCN topological feature (F3) performed relatively weakly when used alone (Macro F1 = 68.3%), because the global topological relationship is difficult to independently support fine style discrimination when lacking local details. The simple concatenation feature (F4) showed a significant improvement compared to the single feature, with a Macro F1 of 82.6%, verifying the complementarity of multi-modal features. However, the dimension of F4 is high (D = 1247), and there is redundant information. The proposed PCA-kernelized combined dimensionality reduction fusion feature (F5) achieved the best performance among all combinations, with a Macro F1 of 88.7% and an accuracy of 89.2%, which is approximately 6 percentage points higher than simple concatenation, proving that the proposed combined dimensionality reduction strategy can effectively remove redundancy and enhance non-linear separability.

To further quantify the significance of the differences between different feature combinations, the performance improvement rate of F5 relative to F4 was calculated:

$$\text{Improvement Rate} = \frac{\text{Macro F1}_{\text{F5}} - \text{Macro F1}_{\text{F4}}}{\text{Macro F1}_{\text{F4}}} \times 100\% = \frac{88.7 - 82.6}{82.6} \times 100\% \approx 7.38\% \quad (22)$$

The above data indicates that the proposed multi-modal feature fusion method has significant performance advantages.

5.2 Comparison of improved algorithm with baseline model

This section conducts a comprehensive comparison of the improved machine learning algorithm proposed in this paper with three baseline models: Support Vector Machine (SVM, RBF kernel, C = 1.0), Random Forest (RF, number of trees T = 200, no limit on maximum depth), and Standard Extreme Learning Machine (ELM, hidden layer nodes L = 200). All models use the same fused feature (F5) as the input to ensure the fairness of the comparison. The evaluation metrics include accuracy, training time, and generalization error. The generalization error is defined as the difference in performance between the training set and the test set of the model, specifically using the absolute value of the difference between the test accuracy and the training accuracy:

$$\epsilon_{\text{gen}} = | \text{ACC}_{\text{train}} - \text{ACC}_{\text{test}} | \quad (23)$$

A smaller ϵ_{gen} indicates that the model has better generalization ability and has not suffered from severe overfitting. The average performance of the four models in 5-fold cross-validation is shown in Table 1.

Table 1. Comparison of performance between improved algorithm and baseline model

Model	Accuracy rate (%)	Training time (seconds)	Generalization error ϵ_{gen} (%)
SVM (RBF)	81.3 ± 1.2	47.6 ± 3.2	3.4 ± 0.5
Random Forest	83.7 ± 1.0	12.3 ± 0.8	2.1 ± 0.3
Standard ELM	86.2 ± 0.9	0.18 ± 0.02	5.8 ± 0.6
Proposed	**91.5 ± 0.7**	0.31 ± 0.03	1.2 ± 0.2

As shown in Table 1, the proposed algorithm achieves an accuracy of 91.5%, significantly higher than SVM (81.3%), Random Forest (83.7%), and Standard ELM (86.2%). The improvement ratio relative to Standard ELM is as follows:

$$\Delta ACC = 91.5\% - 86.2\% = 5.3\% \tag{24}$$

In terms of training time, Standard ELM has the fastest speed because it directly calculates the generalized inverse (0.18 seconds). The improved algorithm in this paper, due to the introduction of adaptive weight calculation and iterative optimization of the interval maximization regularization term, slightly increases the training time to 0.31 seconds, but it is still much lower than SVM (47.6 seconds) and Random Forest (12.3 seconds), and fully meets the practical requirements of engineering efficiency for the EI conference. The most crucial point is the generalization error indicator: Although Standard ELM has a fast training speed, its generalization error is as high as 5.8%, showing a significant tendency of overfitting; while the improved algorithm in this paper has a generalization error of only 1.2%, which is the smallest among all the compared models, thanks to the joint effect of the adaptive margin adjustment mechanism and the interval maximization regularization term. Figure 2 intuitively shows the distribution of the accuracy rates of each fold in the 5-fold cross-validation for the four types of models.

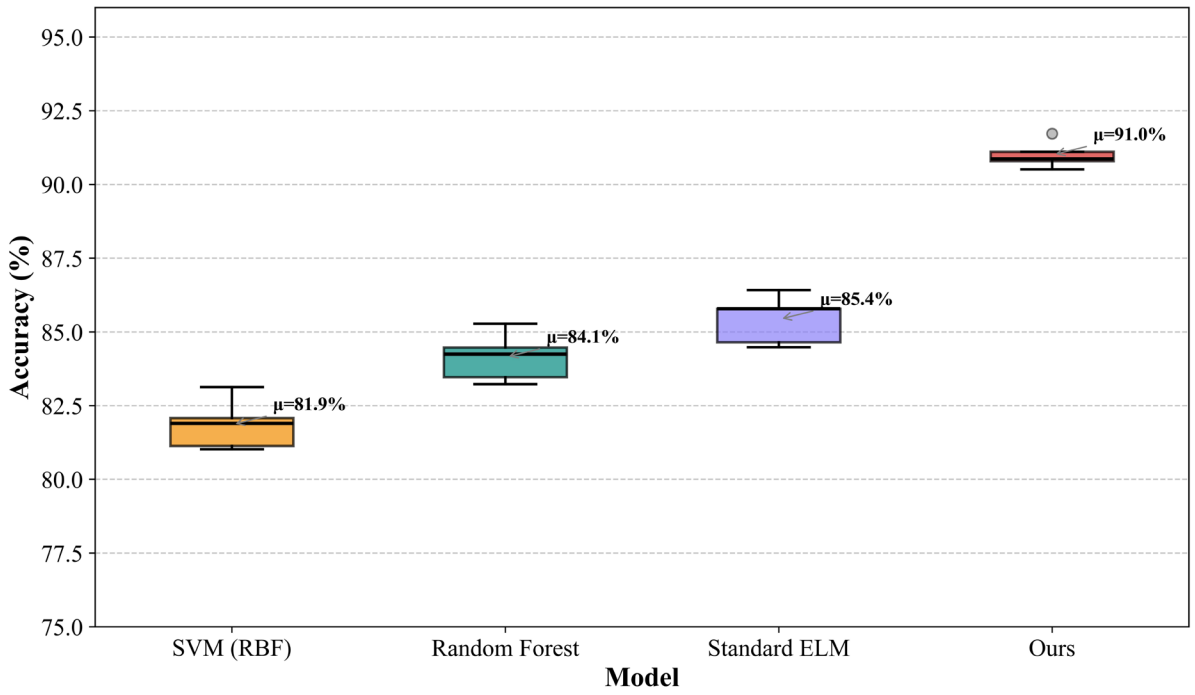

Figure 2. Distribution of accuracy rates of each model in 5-fold cross-validation

The box plot in Figure 2 shows that the box height of the accuracy rate of the improved algorithm in this paper is the narrowest, indicating that its performance fluctuation on the 5-fold test is the smallest, and its stability is superior to other comparison models.

5.3 Ablation experiment

To quantitatively evaluate the individual contributions of each component in the algorithm proposed in this paper, two sets of ablation experiments were designed: feature branch ablation and algorithm module ablation. The feature branch ablation experiment was conducted based on the multi-

branch parallel input architecture (Section 3.3) proposed in this paper, and one feature branch was removed successively to observe the performance changes. The specific settings included: removing the structural contour branch (only retaining texture and topology), removing the texture branch (only retaining contour and topology), removing the topology branch (only retaining contour and texture), and removing both branches (only retaining the single optimal branch). All configurations used the same decision layer weighted voting mechanism. The algorithm module ablation experiment fixedly used the full feature three-branch input and successively removed the core improvement modules of this paper: removing adaptive margin adjustment (using equal-weight ELM), removing architectural innovation (using single-branch feature concatenation instead of three-branch voting), removing the interval maximization regularization term ($\lambda_2 = 0$), and the complete algorithm. The ablation experiment results are shown in Tables 2 and 3.

Table 2. Results of feature branch ablation experiments

Retained feature branches	Macro F1 (%)	Accuracy rate (%)	Relative complete model reduction value
Only outline + texture	84.2	84.7	-4.5
Only outline + topology	82.9	83.3	-5.8
Only texture + topology	79.6	80.1	-9.1
Only outline	74.2	73.8	-14.5
Complete three branches	88.7	89.2	—

As shown in Table 2, removing any of the feature branches leads to a decrease in performance. Among them, when both the contour and texture branches are retained, the performance is relatively better (Macro F1 = 84.2%), while when only the texture and topology branches are retained, the performance drops the most significantly (Macro F1 = 79.6%). This once again verifies the dominant role of structural contour features in style classification. Compared with the single contour branch model, the complete three-branch model has an increase of 14.5 percentage points in Macro F1, fully demonstrating the necessity of multimodal feature fusion.

Table 3. Results of algorithm module abandonment experiment

Configuration	Accuracy rate (%)	Generalization error ϵ_{gen} (%)	Training time (seconds)
Remove adaptive margin adjustment	87.3	3.8	0.21
Remove architectural innovation (single branch)	86.8	2.5	0.22
Remove the interval maximization regularization term ($\lambda_2 = 0$)	88.6	2.1	0.27
Complete algorithm	91.5	1.2	0.31

The algorithm module ablation results in Table 3 show that each of the three improved modules has a positive contribution to the overall performance. Among them, removing the adaptive margin adjustment led to the greatest decrease in accuracy (from 91.5% to 87.3%, a decrease of 4.2 percentage

points), while the generalization error increased from 1.2% to 3.8%, indicating that the weighted processing of boundary samples by this module is the key to suppressing overfitting. Removing the architectural innovation (reducing to a single-branch feature concatenation) caused the accuracy to drop to 86.8%, verifying the effectiveness of the multi-branch parallel input and decision layer weighted voting mechanism. After removing the maximum margin regularization term, the accuracy dropped to 88.6% and the generalization error increased to 2.1%, indicating that this regularization term indeed enhanced the discriminative ability by widening the margins between categories. The combined use of the three modules brought synergistic benefits, and the accuracy of the complete algorithm was higher than the configuration after removing any single module.

5.4 Robustness Test

To evaluate the robustness of the model in real engineering applications, two types of perturbation experiments were designed: noise injection and missing feature. In the noise injection experiment, Gaussian noise $\xi \sim \mathcal{N}(0, \sigma^2 I)$ was added to the input feature vector x, where σ is the noise standard deviation and I is the identity matrix. The signal-to-noise ratio (SNR) is defined as:

$$\mathrm{SNR} = 10 \log_{10}\left(\frac{\| x \|_2^2}{\| \xi \|_2^2}\right) (\mathrm{dB}) \tag{25}$$

Four noise levels were set in the experiment: SNR = 30 dB (mild noise), 20 dB (moderate noise), 10 dB (severe noise), and 5 dB (extreme noise). In the missing feature experiment, a certain part of the input feature vector was randomly set to zero to simulate sensor failure or feature extraction failure, and three missing proportions were set: 10%, 20%, and 30%. The missing positions were randomly generated in each cross-validation to ensure statistical reliability. The robustness test results are shown in Table 4.

Table 4. Results of robustness test

Noise level (SNR)	Proposed	Standard ELM	Random Forest
No noise	91.5	86.2	83.7
30 dB	90.8	85.1	82.9
20 dB	88.3	81.5	79.4
10 dB	82.1	72.3	68.7
5 dB	74.6	60.8	56.2

From the noise experiment section of Table 4, it can be seen that as the noise intensity increases (SNR decreasing from 30 dB to 5 dB), the accuracy of all models shows a downward trend. The algorithm in this paper still maintains an accuracy rate of 82.1% at SNR = 10 dB, while the standard ELM and random forest drop to 72.3% and 68.7% respectively. Under extremely heavy noise (SNR = 5 dB) conditions, the accuracy of this algorithm is 74.6%, which is 13.8 percentage points higher than that of the standard ELM (60.8%). The noise robustness coefficient is defined as:

$$\rho_{\mathrm{noise}} = \frac{\mathrm{ACC}_{\mathrm{SNR=5dB}}}{\mathrm{ACC}_{\mathrm{no\ noise}}} \times 100\% \tag{26}$$

The $\rho_{\text{noise}} = 74.6/91.5 \times 100\% \approx 81.5\%$, while the standard ELM is only $60.8/86.2 \times 100\% \approx 70.5\%$. This indicates that the algorithm in this paper has stronger tolerance to noise. This is mainly attributed to the adaptive margin adjustment mechanism that gives higher weight to boundary samples during training, enabling the model to have better robustness to feature perturbations. As shown in Table 5, under different missing proportions, the accuracy of this algorithm is significantly better than that of the standard ELM and random forest, and as the missing proportion increases, the performance of this algorithm decays most slowly, demonstrating stronger tolerance to missing values.

Table 5. Comparison of accuracy rates of three algorithms under different missing ratio conditions

Proportion of absence	The algorithm of this article	Standard ELM	Random Forest
0%	91.5	86.2	83.7
10%	89.2	82.4	79.8
20%	85.7	76.1	72.5
30%	79.4	67.3	63.1

From the missing feature experiment section, it can be seen that when 30% of the features are missing, the accuracy of the algorithm in this paper is 79.4%, a decrease of 12.1 percentage points; while the standard ELM decreases by 18.9 percentage points, and the random forest decreases by 20.6 percentage points. Define the missing tolerance index:

$$\tau_{\text{missing}} = \text{ACC}_{30\%\ \text{missing}} - \text{ACC}_{\text{others}} \tag{27}$$

The absolute performance of the algorithm in this paper under 30% missing (79.4%) is still higher than the performance of the standard ELM under 20% missing (76.1%), indicating that the algorithm in this paper has a better redundancy tolerance for feature missing. This is due to the multi-branch parallel architecture - even if some feature branches are severely disturbed, other branches can still provide effective information through the weighted voting mechanism of the decision layer. In summary, the experiments in this Section comprehensively verify the effectiveness and engineering applicability of the proposed multimodal feature extraction and improved classification algorithm through feature comparison, model comparison, ablation analysis, and robustness testing.

6. CONCLUSION AND FUTURE WORK

This paper addresses the issue of automatic extraction and classification of design scheme style characteristics, proposing a systematic method that integrates multimodal feature extraction and an improved integrated classification architecture. At the feature extraction level, it breaks through the limitations of traditional manual design features, constructing a multimodal feature space from three dimensions: structural contour, local texture, and global topological relationship. Specifically, the enhanced HOG features with adaptive weighting based on gradient amplitude have strengthened the discriminative information of contour edges, the LBP-variant operator has been designed to effectively capture the periodic repetitive texture patterns in the style, and the graph convolution network has been introduced to embed the global topological relationships between design elements. On this basis, the

PCA-kernelized joint dimension reduction strategy is adopted to remove feature redundancy while retaining the nonlinear discriminant structure, providing information-complete and dimensionally-controllable feature inputs for the subsequent classification. At the classification algorithm level, in response to the problems of low processing efficiency, poor boundary sample sensitivity, and insufficient category intervals in existing baseline models on multimodal fusion features, this paper proposes an adaptive margin adjustment weighted extreme learning machine as the core classifier, designs an innovative architecture with parallel input of multiple branches and weighted voting at the decision layer, and reconstructs the loss function by introducing a regularization term based on the maximum style interval. Experimental results show that the proposed fused features achieve a Macro F1 score of 88.7%, an improvement of approximately 7.4 percentage points compared to simple feature concatenation; the complete improved algorithm achieves an accuracy of 91.5%, an increase of 5.3 percentage points compared to standard ELM, and significantly reduces the generalization error from 5.8% to 1.2%. Abandonment experiments verify the effectiveness of the three core improvement modules and their synergistic gain, and robustness tests further prove that the algorithm can maintain high classification performance under noise interference and feature loss conditions. From an engineering application value perspective, this method is oriented towards the practicality and reproducibility requirements of the EI conference, and the overall process has clear modular characteristics. The multimodal feature extraction process does not rely on large-scale pre-trained models, and the computational cost is controllable; the training time of the weighted extreme learning machine is only 0.31 seconds, much lower than support vector machines and random forests, meeting the requirements of rapid iteration in engineering scenarios. At the same time, the introduction of hierarchical k-fold cross-validation and statistical significance tests ensures the reliability and reproducibility of the experimental results. This method can be extended to practical application scenarios such as architectural design scheme style retrieval, automatic annotation of interior decoration styles, and classification of industrial product shapes, providing technical support for the efficient organization and intelligent management of design resources.

In future work, several directions are worthy of further exploration for this method. First, the hyperparameters such as the gradient enhancement factor and LBP variation vector in the current feature extraction process are mainly set based on experience. In the future, a meta-learning or adaptive parameter prediction mechanism can be introduced to enable the feature extraction process to dynamically adjust according to the statistical characteristics of the input design scheme, further improving the model's adaptability to different style domains. Second, although this paper has compressed the feature dimensions to some extent through PCA-kernelized joint dimension reduction, the extraction of the three modal features is still relatively independent. In the future, an end-to-end lightweight neural network architecture can be designed to unify the learning of multimodal features and the training of the classifier within the same optimization framework, and explore model pruning and quantization techniques to reduce the storage and computational costs during deployment, facilitating real-time operation on mobile devices or embedded systems. Third, the current research focuses on the classification task of eight typical design styles. In the future, it can be expanded to fine-grained style subclass recognition and feature invariance learning in cross-domain style transfer scenarios, further expanding the generalization boundary of the method.

Abbreviations

HOG, Histogram of Oriented Gradients;

LBP, Local Binary Pattern;

GCN, Graph Convolutional Network;

PCA, Principal Component Analysis;

ELM, Extreme Learning Machine;

SVM, Support Vector Machine;

RF, Random Forest;

RBF, Radial Basis Function;

ReLU, Rectified Linear Unit;

TP, True Positive;

TN, True Negative;

FP, False Positive;

FN, False Negative;

SNR, Signal-to-Noise Ratio;

dB, Decibel;

F1, F1-Score;

Macro F1, Macro Average F1-Score;

ACC, Accuracy;

GFLOPs, Giga Floating Point Operations per Second;

L-BFGS, Limited-memory Broyden–Fletcher–Goldfarb–Shanno;

CPU, Central Processing Unit;

GPU, Graphics Processing Unit.

Supplementary Material

Not applicable.

Appendix

Not applicable.

Ethics approval and consent to participate.

This study did not involve human participants, animal subjects, or any data requiring ethical approval. Therefore, ethics approval and consent to participate are not applicable.

Acknowledgements

The authors would like to thank the editors of this journal and all the anonymous reviewers who provided valuable comments on this work.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Author contributions

All authors have read and agreed to the published version of the manuscript. The author's contributions are specified as follows: **W.W.:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original draft, Writing – Review & Editing, Visualization, Supervision, Project administration.

Funding information

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The data that support the findings of this study are available upon request from the corresponding authors, **W.W.**

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's results, findings, and content.

Declaration of AI and AI-assisted Technologies in the Writing Process

During the writing of this article, the author used DeepSeek for spelling and grammar checking. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the final published content.

REFERENCES

[1] Liu, X., & Zhang, C. (2024). Data Mining and Pattern Recognition of the Integration of Art Design Style and CAD System. *Computer-Aided Design and Applications*, *21*, 116-131. https://doi.org/10.14733/cadaps.2024.s19.116-131

[2] Xin, T., Kim, S., & Shangshang, Z. (2025). Classification and Analysis of Packaging Design Styles Based on Visual Feature Extraction. *IEEE Access*, *13*, 206639-206653. https://doi.org/10.1109/access.2025.3634619

[3] Vijendran, M., Deng, J., Chen, S., Ho, E. S., & Shum, H. P. (2024). Artificial intelligence for geometry-based feature extraction, analysis and synthesis in artistic images: a survey. *arXiv preprint arXiv:2412.01450*. https://doi.org/10.1007/s10462-024-11051-3

[4] Kumar, D., Pandey, R. C., & Mishra, A. K. (2024). A review of image features extraction techniques and their applications in image forensic. *Multimedia tools and applications*, *83*(40), 87801-87902. https://doi.org/10.1007/s11042-023-17950-x

[5] Rundo, L., & Militello, C. (2024). Image biomarkers and explainable AI: handcrafted features versus deep learned features. *European Radiology Experimental*, *8*(1), 130. https://doi.org/10.1186/s41747-024-00529-y

[6] He, J., Guo, J., Zhou, L., & Liu, Y. (2025). Hybrid deep learning and fractional Brownian motion approach for probabilistic RUL prediction in industrial equipment. *IEEE Transactions on Instrumentation and Measurement*. https://doi.org/10.1109/tim.2025.3606063

[7] Feng, H., Zhong, C., Zhang, Z., Liu, Y., & Ma, Q. (2026). FFAKAN: A Frequency-Aware Filtering Activation-Based Kolmogorov-Arnold Network for Hyperspectral Image Classification. *Remote Sensing*, *18*(7), 981. https://doi.org/10.3390/rs18070981

[8] Islam, M. R., Lima, A. A., Das, S. C., Mridha, M. F., Prodeep, A. R., & Watanobe, Y. (2022). A comprehensive survey on the process, methods, evaluation, and challenges of feature selection. *IEEE access*, *10*, 99595-99632. https://doi.org/10.1109/access.2022.3205618

[9] Xin, T., Kim, S., & Shangshang, Z. (2025). Classification and Analysis of Packaging Design Styles Based on Visual Feature Extraction. *IEEE Access*, *13*, 206639-206653. https://doi.org/10.1109/access.2025.3634619

[10] Peng, D., Liu, R., Lu, J., & Zhang, S. (2022). Unsupervised multi-modal modeling of fashion styles with visual attributes. *Applied Soft Computing*, *115*, 108214. https://doi.org/10.1016/j.asoc.2021.108214

[11] Wulandari, M., Chai, R., Basari, B., & Gunawan, D. (2024). Hybrid feature extractor using discrete wavelet transform and histogram of oriented gradient on Convolutional-Neural-Network-Based palm vein recognition. *Sensors*, *24*(2), 341. https://doi.org/10.3390/s24020341

[12] Bouchene, M. M. (2024). Bayesian optimization of histogram of oriented gradients (HOG) parameters for facial recognition: MM Bouchene. *The Journal of Supercomputing*, *80*(14), 20118-20149. https://doi.org/10.1007/s11227-024-06259-7

[13] Bakheet, S. (2017). An SVM framework for malignant melanoma detection based on optimized HOG features. *Computation*, *5*(1), 4. https://doi.org/10.3390/computation5010004

[14] Xu, X., & Li, B. (2025). Multi scale supervised entropy weighted binary pattern for texture

classification. *Scientific Reports*, *15*(1), 26087. https://doi.org/10.1038/s41598-025-11245-x

[15] Maher, H. (2024). Texture Analysis and Classification using Local Binary Patterns and Statistical Features. *Wasit Journal of Computer and Mathematics Science*, *3*(3), 79-88. https://doi.org/10.31185/wjcms.279

[16] Elazab, N., Gab Allah, W., & Elmogy, M. (2024). Computer-aided diagnosis system for grading brain tumor using histopathology images based on color and texture features. *BMC Medical Imaging*, *24*(1), 177. https://doi.org/10.1186/s12880-024-01355-9

[17] Kaur, M. (2023). A comparative analysis of local binary pattern (LBP) variants for image tamper detection. https://doi.org/10.21203/rs.3.rs-3608580/v1

[18] Al Saidi, I., Rziza, M., & Debayle, J. (2023). Completed homogeneous LBP for remote sensing image classification. *International Journal of Remote Sensing*, *44*(12), 3815-3836. https://doi.org/10.1080/01431161.2023.2227320

[19] Meeradevi, T., Sasikala, S., Gomathi, S., & Prabakaran, K. (2023). An analytical survey of textile fabric defect and shade variation detection system using image processing. *Multimedia Tools and Applications*, *82*(4), 6167-6196. https://doi.org/10.1007/s11042-022-13575-8

[20] Lu, J., Zhang, C., Li, J., Zhao, Y., Qiu, W., Li, T., ... & He, J. (2022). Graph convolutional networks-based method for estimating design loads of complex buildings in the preliminary design stage. *Applied Energy*, *322*, 119478. https://doi.org/10.1016/j.apenergy.2022.119478

[21] Zhou, Y., Huo, H., Hou, Z., Bu, L., Mao, J., Wang, Y., ... & Bu, F. (2023). Co-embedding of edges and nodes with deep graph convolutional neural networks. *Scientific Reports*, *13*(1), 16966. https://doi.org/10.1038/s41598-023-44224-1

[22] Baranwal, A., Fountoulakis, K., & Jagannath, A. (2022). Effects of graph convolutions in multi-layer networks. *arXiv preprint arXiv:2204.09297*. https://doi.org/10.48550/arXiv.2204.09297

[23] Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*(1), 100. https://doi.org/10.1038/s43586-023-00209-y

[24] Bharadiya, J. P. (2023). A tutorial on principal component analysis for dimensionality reduction in machine learning. *International journal of innovative science and research technology*, *8*(5), 2028-2032. https://doi.org/10.5281/zenodo.8002436

[25] Ashraf, A., Nawi, N. M., & Aamir, M. (2023). Adaptive feature selection and image classification using manifold learning techniques. *IEEE Access*, *12*, 40279-40289. https://doi.org/10.1109/access.2023.3322147

[26] Jia, W., Sun, M., Lian, J., & Hou, S. (2022). Feature dimensionality reduction: a review. *Complex & Intelligent Systems*, *8*(3), 2663-2693. https://doi.org/10.1007/s40747-021-00637-x

[27] Song, W., Zhang, X., Yang, G., Chen, Y., Wang, L., & Xu, H. (2024). A study on dimensionality reduction and parameters for hyperspectral imagery based on manifold learning. *Sensors*, *24*(7), 2089. https://doi.org/10.3390/s24072089

[28] Jiao, T., Guo, C., Feng, X., Chen, Y., & Song, J. (2024). A comprehensive survey on deep learning multi-modal fusion: Methods, technologies and applications. *Computers, Materials & Continua*, *80*(1).

https://doi.org/10.32604/cmc.2024.053204

[29] Gao, X., Zhang, G., & Xiong, Y. (2022). Multi-scale multi-modal fusion for object detection in autonomous driving based on selective kernel. *Measurement*, *194*, 111001. https://doi.org/10.1109/TVT.2022.3230265

[30] Kullu, O., & Cinar, E. (2022). A deep-learning-based multi-modal sensor fusion approach for detection of equipment faults. *Machines*, *10*(11), 1105. https://doi.org/10.3390/machines10111105

[31] Lu, T., Xiong, J., Zhou, J., Wang, Q., Cen, J., Liu, M., ... & Zhang, J. (2024). Bearing fault diagnosis using multichannel broad learning system based on positive–negative weighted voting mechanism. *IEEE Transactions on Instrumentation and Measurement*, *73*, 1-10. https://doi.org/10.1109/TIM.2024.3373075

[32] Khan, A., Ali, A., Islam, N., Manzoor, S., Zeb, H., Azeem, M., & Ihtesham, S. (2022). Robust Extreme Learning Machine Using New Activation and Loss Functions Based on M-Estimation for Regression and Classification. *Scientific programming*, *2022*(1), 6446080. https://doi.org/10.1155/2022/6446080

[33] Atallah, E. S., Hagag, A., Ali, E. M., & Abassy, T. A. (2025). Using the Quasi-Newton Method to Solve Nonlinear Least Squares Regression Problems. *International Journal of Applied Intelligent Computing and Informatics*, *1*(1), 9-15. https://doi.org/10.21608/ijaici.2025.340085.1004

APA Citation

Wang, W. (2026). Design Scheme Style Feature Extraction and Machine Learning Classification Algorithm Construction. *International Scientific Technical and Economic Research, 4*(3), 1-21. **https://doi.org/10.67541/istaer2623**